

文章编号: 1000-5641(2017)05-0066-14

# 基于卷积神经网络的自动问答

金丽娇, 傅云斌, 董启文

(华东师范大学 数据科学与工程学院, 上海 200062)

**摘要:** 自动问答是自然语言处理领域中的一个研究热点, 自动问答系统能够用简短、精确的答案直接回答用户提出的问题, 给用户提供更加精确的信息服务. 自动问答系统中需解决两个关键问题: 一是实现自然语言问句及答案的语义表示, 另一个是实现问句及答案间的语义匹配. 卷积神经网络是一种经典的深层网络结构, 近年来卷积神经网络在自然语言处理领域表现出强大的语言表示能力, 被广泛应用于自动问答领域中. 本文对基于卷积神经网络的自动问答技术进行了梳理和总结, 从语义表示和语义匹配两个主要角度分别对面向知识库和面向文本的问答技术进行了归纳, 并指出了当前的研究难点.

**关键词:** 卷积神经网络; 自动问答; 语义表示; 语义匹配

**中图分类号:** TP391    **文献标志码:** A    **DOI:** 10.3969/j.issn.1000-5641.2017.05.007

## The auto-question answering system based on convolution neural network

JING Li-jiao, FU Yun-bin, DONG Qi-wen

(School of Data Science and Engineering, East China Normal University, Shanghai 200062, China)

**Abstract:** The question-answering is a hot research field in natural language processing, which can give users concise and precise answer to the question presented in natural language and provide the users with more accurate information service. There are two key questions to be solved in the question answering system: one is to realize the semantic representation of natural language question and answer, and the other is to realize the semantic matching learning between question and answer. Convolution neural network is a classic deep network structure which has a strong ability to express semantics in the field of natural language processing in recent years, and is widely used in the field of automatic question and answer. This paper reviews some techniques in the question answering system that is based on the convolution neural network, the paper focuses on the knowledge-based and the text-oriented Q&A techniques from the two main perspectives

收稿日期: 2017-06-23

基金项目: 国家重点研发计划(2016YFB1000905); 国家自然科学基金广东省联合重点项目(U1401256);  
国家自然科学基金(61672234, 61402177); 华东师范大学信息化软课题

第一作者: 金丽娇, 女, 硕士研究生, 研究方向为自然语言处理与自动问答.

E-mail: 51164500102@stu.ecnu.edu.cn.

通信作者: 傅云斌, 男, 博士后, 研究方向为数据科学与机器学习. E-mail: fuyunbin2012@163.com.

of semantic representation and semantic matching, and indicates the current research difficulties.

**Key words:** convolution neural network; automatic question answering; semantic representation; semantic matching

## 0 引言

随着信息技术的发展,人们对信息的精准化需求也越来越高,传统的搜索引擎主要是基于关键字匹配的检索模式,然后返回相关文档.然而在很多情况下,用户可能只是想知道某一问题的具体答案,而不是一堆文档.自动问答系统允许用户以自然语言形式提出问题,并采用自然语言处理技术自动地将简洁、正确的回答返回给用户.

早在上世纪60年代人工智能刚刚兴起的时候,人们就提出过自动问答的相关概念,但由于当时技术和条件的限制,问答系统并没有得到很大的发展.随着近几年来信息检索、自然语言处理和人工智能等相关领域的发展,为自动问答系统的构建创造了新的条件.

Start 系统<sup>[1]</sup>是最早的自动问答系统,1993年由麻省理工学院 Boris Katz 等人开发,可以回答有关电影、科技、历史等方面的简单问题.除了Start系统,askJeeves<sup>[2]</sup>、AnswerBus<sup>[3]</sup>等系统都是目前较为成熟的问答系统,但它们返回包含结果的句子或段落.

目前的搜索引擎大多基于信息检索技术,即先对问句进行简单的分析,提取关键字,然后根据关键字从相应的资源库中查询并返回一系列相关文档或网页.而自动问答系统包含信息抽取、自然语言处理等多方面的研究技术.一个问答系统一般由问句分析、信息检索和答案抽取三部分组成<sup>[4]</sup>.目前基于卷积神经网络的问答任务实现过程主要是先利用信息检索技术对海量文本进行过滤,再将自然语言问句和答案集合通过一定的技术匹配,最后进行相似度计算来抽取答案.自动问答任务需要对问句进行更细致的分析,处理更复杂的语法、语义分析,问句和相关文档需要进行更精细的匹配.

根据不同的分类标准,问答系统可以划分成多种类型.根据获取问题答案所依存的信息资源的不同,问答系统可以被粗略地分为基于知识库的问答系统和基于文本的问答系统.基于知识库的问答系统面向的是结构化的数据,基于文本的问答系统面向的一般是自然语言句子.根据信息资源库数据不同的表现形式,问答系统的实现技术也千差万别.文献<sup>[5]</sup>从不同的实现技术角度,将问答技术分为两大类型,一类基于语义解析,另一类则基于信息抽取.

基于语义解析的问答系统的任务是将自然语言问句转化为结构化、计算机可理解的语义表示形式,例如 $\lambda$ 范式<sup>[6]</sup>、DCS-Tree<sup>[7]</sup>等,然后在相应的知识库上查询,推理出匹配的答案.现有的知识库包括维基百科等,基于语义解析的问答系统一般都是面向知识库的.而基于信息抽取的问答系统则需从大量资源中先查找候选文档,再根据问句语义信息等特征,从这些候选文档中抽取答案.

近年来,由于深度学习在多个领域都取得了突破性进展,很多研究人员开始研究深度学习模型在自然语言处理上的应用.深度学习模型为实现句子和文档的语义表示提供了新的方案,特别是用基于神经网络的方法来进行问句和答案句的语义表示、为基于知识库技术的问答系统提供了新的解决思路,同时在语义匹配上,也相比传统的方法有了更好的效果.本文从语义表示和语义匹配的角度出发,并重点关注深度学习中的一个分支——卷积神经网络在自动问答领域中的应用.

本文内容组织如下. 第 1 节对基于卷积神经网络的语义表示做简要介绍; 第 2 节对语义匹配技术以及基于卷积神经网络的现有模型进行概述; 第 3 节从语义表示学习和语义匹配模型两方面出发, 分别描述利用卷积神经网络, 在基于知识库和基于文本的问答系统中的实现思路; 最后, 第 4 节对本文进行了简单总结.

## 1 卷积神经网络的语义表示

### 1.1 概述

传统的文本表示方法是采用词袋表示模型<sup>[8]</sup>, 它用向量来实现文本的表示. 在词袋模型中, 常用的特征有二元词组、 $n$ 元词组或一些人工抽取的特征, 用词袋表示上下文信息的词表示模型有基于矩阵的 LSA 模型<sup>[9]</sup>、Brown 等人提出的 clustering 模型<sup>[10]</sup>. 但是这种向量是高维稀疏的, 而且忽略了词序, 因此不能有效地表示句子的上下文信息. 后来, Bengio 等人提出神经语言模型, 同时在学习时得到了一种名为词嵌入的向量表示方法<sup>[11]</sup>, 词嵌入表示技术通过神经网络技术对上下文、上下文和目标词的关系进行建模, 是一种低维、稠密、连续的向量表示, 可以表示复杂的上下文. 并且, 词嵌入可以从大量无标注语料中自动学习句法和语义特征<sup>[12]</sup>, 现有的词嵌入模型有 skip-gram, CBOW<sup>[13]</sup>, C&W<sup>[14]</sup>, LBL<sup>[15]</sup>, NNLM<sup>[11]</sup>等. 目前, 基于深度学习的自动问答系统, 一般都是基于词嵌入的表示.

然而, 在实际的需求中, 如果仅仅使用词级别的语义表示, 还远远不能满足现实场景, 句子和篇章级别的语义表示相对来说意义会更大. 关于句子和文档的语义表示, 德国数学家弗雷格曾指出: 一段话的语义由其各组合部分的语义以及它们之间的组成方法所确定<sup>[16]</sup>, 进而提出组合语义的思想. 常用的组合语义函数有加权组合、矩阵乘法、张量乘法等<sup>[17]</sup>. 从神经网络角度上看, 则主要可分为三类: 循环神经网络, 递归神经网络和卷积神经网络<sup>[18]</sup>. 本文主要对卷积神经网络进行展开描述.

### 1.2 卷积神经网络的语义表示模型

卷积神经网络(CNN)最早由 Fukushima 在 1980 年提出<sup>[19]</sup>, 此后, LeCun 等人对其做了重要改进<sup>[20]</sup>, 结构如图 1<sup>[21]</sup>所示. 其核心思想是局部感知和权值共享, 基本结构主要包括两层, 一层为卷积层, 又可称特征提取层, 另一层为池化层, 又称为特征映射层.

在卷积层, 通过一个固定窗口的卷积核(滤波器)对文本中每个部分的局部信息进行建模, 一般一个卷积层会有多个卷积核. 输入层通过多个卷积核的卷积操作会得到若干个特征映射面, 卷积核是一个权值矩阵, 对于每一个卷积核, 对句子从左到右进行扫描, 窗口内有多多个单词, 每个单词由一个向量表示, 卷积核与窗口内的多个词向量进行局部连接. 然后该局部加权和传递给一个激活函数, 最后的输出为一个特征映射向量, 对于同一特征映射面, 每一个节点都是权值共享的. 对于输入的句子  $S$ , 用一个窗口为  $m$  的卷积核  $W$  作用后, 得到一个向量  $h_i$ , 具体公式为

$$h_i = (S * W)_i = S_{[i:i+m-1]}^T \cdot W,$$

然后  $h_i$  再通过一个激活函数, 得到一个特征映射向量  $c_i$ :

$$c_i = \alpha(h_i + b_i),$$

其中  $\alpha$  为非线性激活函数, 一般有 sigmoid 函数, tanh 函数和 Relu 等<sup>[22]</sup>,  $b_i$  为偏置项.

池化是对所有的语义组合进行选择, 即从每一个特征映射面中选择固定数目的节点, 最后将句子表示为一个固定长度的向量. 常用的池化技术有均值池化和最大值池化<sup>[23]</sup>, 池化操作不仅可以解决不同长度句子的输入问题, 同时可以起到降维的作用.

在卷积神经网络中, 一般会有多个卷积层和池化层, 经过一个或若干个卷积层和池化层后, 最终会得到表示句子语义信息的输出向量. 随着深度的加深, 模型得到的输出表示可以捕捉全局语义信息和更加高级的特征.

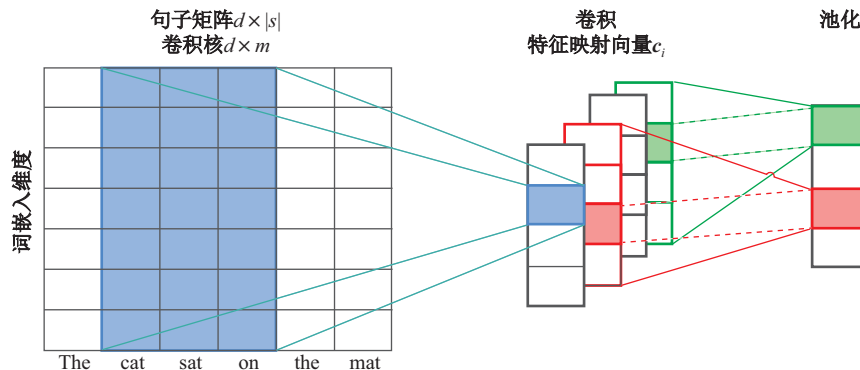


图1 卷积网络结构图

Fig. 1 The architecture of convolution network

## 2 基于卷积神经网络的语义匹配

### 2.1 概述

语义匹配是自动问答的核心技术, 在自动问答系统中, 问句分析模块(问题的复述)、信息检索模块和答案的抽取模块都主要涉及到语义匹配技术. 语义匹配基于语义表示, 从不同的粒度考虑, 语义匹配有词、短语、句子、文本等级别的匹配.

传统的文本语义匹配模型, 如 Berger 等人提出的翻译模型<sup>[24]</sup>, Wu 等人提出的正则化隐空间映射<sup>[25]</sup>等, 往往需要依赖于人工定制的特征和外部的语义资源库<sup>[26-27]</sup>, 这种基于手工定义特征的模型大部分都针对于一些特定的问题的应用(如答案抽取, 信息检索), 没有很好的扩展性. 因此, 传统的语义匹配方法存在效率低, 扩展性差, 无法挖掘大量数据中的隐含特征等问题. 利用深度学习技术, 可以在没有人工干预的情况下在大量样本中自动学习选择特征, 省去了手工定义特征的开销. 除此之外, 通过结合词向量技术(如 Word2Vec<sup>[28]</sup>)以及神经网络的层次化设计, 可以解决词语匹配的多元性、短语匹配的结构性和文本匹配的层次性问题<sup>[29]</sup>.

### 2.2 基于卷积神经网络的语义匹配模型

本文将重点分析基于卷积神经网络的语义匹配模型, 并根据不同的应用场景, 将问答领域的匹配模型分为两类: 并列式匹配模型(如图2所示)和交互式匹配模型(如图3所示). 并列式匹配模型的思路是, 先将两个句子分别通过卷积网络表示成一个包含语义信息的稠密向量, 然后直接计算两个向量的相似度<sup>[30]</sup>, 或者再通过一个简单的神经网络输出两个句子的匹配分数. 如果是处理比较复杂的问答语句, 则需要从多个角度处理问答句的语义信息<sup>[31]</sup>. 并列式匹配模型注重于文本的表示, 交互式模型则是直接捕获问句和答案句之间的匹配特征, 将问句与答案句同时送入一个卷积网络, 最后学习得到的是问答句的匹配特征, 从而更

加注意匹配的本质.

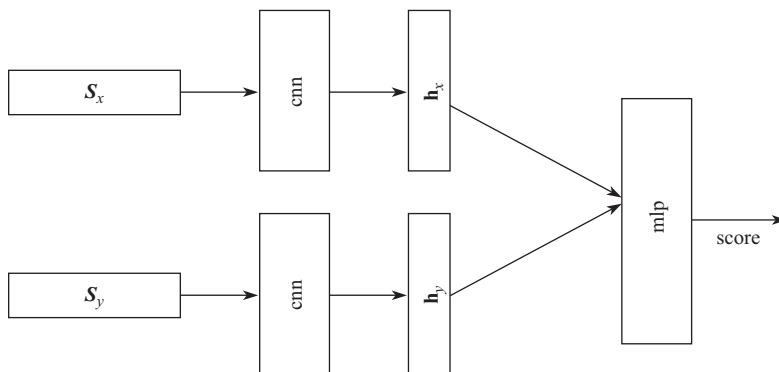


图2 并列式匹配模型

Fig. 2 Parallel matching model

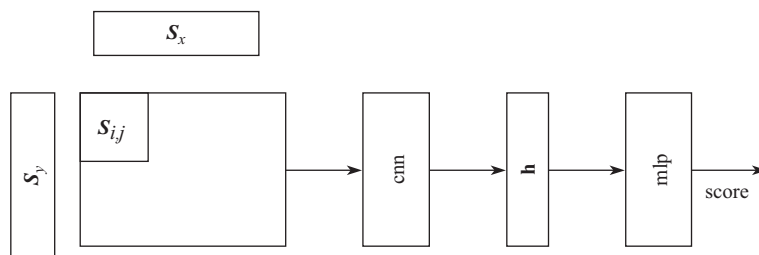


图3 交互式匹配模型

Fig. 3 Interactive matching model

### 3 基于卷积神经网络的自动问答

#### 3.1 基于卷积神经网络的知识库问答

知识库中的数据以三元组(e1,p,e2)的形式保存, e1, e2 表示实体对象, p 表示实体的谓语关系. 目前, 基于知识库的问答系统一般有两种实现方式, 其一是传统的基于语义解析的方法, 这种方法主要是先将句子的主谓宾等成分抽取出来, 再转化成结构化的表达形式, 最终推理出正确的答案, 如组合范畴语法(CCG)<sup>[32]</sup>, 依存语义分析<sup>[7]</sup>等, 然而构建这样的解析器需要大量的人工标注, 代价昂贵, 而且效果并不是很好. 另外一种方法就是基于深度学习模型, 该方法的核心就是把自然语言问句和知识库中的数据都通过深度学习模型映射到一个低维的向量空间, 这样问句和以结构化形式表示的答案都可以用向量来表示, 基于知识库的问答就转化成求解向量相似度的问题, 该方法最先由 Bordes 等人<sup>[33]</sup>提出, 他们的方法在 Reverb 数据集上取得了不错的实验效果, 不过对问句和知识库的语义分析仅仅是简单的对基于实体、关系向量的求和. 同年, Bordes 等人在文<sup>[33]</sup>的基础上进行了改进, 提出子图嵌入的表示方法<sup>[34]</sup>, 他们从3种角度来表示答案: 1. 单个实体的向量表示; 2. 答案路径的向量表示; 3. 子图的向量表示, 即与答案直接关联的实体和路径的向量表示. 该模型在 WEBQUESTIONS 上的实验效果要比之前的模型好. Bordes 等人原先模型<sup>[33]</sup>的 F1 值为 29.7%, 改进后模型<sup>[34]</sup>的 F1 值为 39.2%.

在 Huang 等人<sup>[35]</sup>, Shen 等人<sup>[36]</sup>对问答句中语义距离分析的研究基础上, 针对单语义的问句, Yih 等人<sup>[30]</sup>从实体和关系两个层面进行匹配, 将问句拆分为实体集合和关系集合, 在

对自然语言问句的语义表示学习上, 采用字母三元组的计数向量对词向量进行初始化, 并使用深层卷积网络进行句子的语义学习. 在语义匹配模型上, 实现问句和关系与知识库中的实体和关系分别映射. 作者采用枚举的方法, 列出句子所有可能的分解组合, 其中实体集合为问题中连续词语的子序列, 关系模板为问句实体被特殊符号代替后的对应句子, 然后将关系模板与知识库的所有关系词进行映射匹配, 得到候选的关系集, 接着找出所有的包含候选关系集的三元组, 再进行实体的映射匹配, 找到对应的实体和关系后, 就可以从知识库中找到其指向的答案实体. 参考 Collobert 等人<sup>[23]</sup>的卷积神经网络架构, Yih 等人提出了基于语义模型的卷积神经网络结构(CNNSM)来处理自然语言句子. CNNSM 模型的结构如图 4<sup>[30]</sup>所示, 该模型对输入的问句实体(关系)先进行词哈希操作<sup>[35]</sup>, 初始词向量的表示采用“字母三元组”计数向量, 然后进行卷积、最大值池化操作, 池化后的输出向量与一个语义预测矩阵相乘后经过激活函数处理, 最后输出一个表示语义层面的向量. 相应地, 知识库中的实体或关系也用向量表示, 训练时, 分别训练实体和关系两个模型, 但方法基本类似. 以关系匹配为例, 其目标函数为

$$p(R^+|Q) = \frac{\exp(\gamma \cdot \cos(\mathbf{y}_{R^+}, \mathbf{y}_Q))}{\sum_{R'} \exp(\gamma \cdot \cos(\mathbf{y}_{R'}, \mathbf{y}_Q))},$$

其中  $\mathbf{y}_Q$  为问句的关系模板的输出向量,  $\mathbf{y}_{R^+}$  为知识库中与问句匹配度高的关系向量表示,  $\gamma$  是一个缩放因子.

该实验的训练数据来自 Paralex<sup>[37]</sup>, 其答案的数据来自 ReVerb<sup>[38]</sup>数据集, 数据以三元组的形式保存, 但并不是完全结构化的数据. Yih 等人分别进行了两次实验, 实验结果表明, 考虑实体和关系两个层面匹配的结果要比只考虑实体匹配的结果好, F1 值分别为 0.57 和 0.54, Yih 等人将实验结果与 Paralex 系统进行了对比, Paralex 系统<sup>[37]</sup>也是从句子的关系和实体两方面考虑, 但该模型的目的是建立一个语义词典, 这种基于语义模型的卷积神经网络结构要好于 Paralex.

文献 [30] 提出的这种模型是针对单语义问答, 比较注重句子的语义表示学习, 匹配模式相对简单. 2015 年, Yih 等人受 Yao 等人<sup>[39]</sup>、Bao 等人<sup>[40]</sup>的启发, 针对大型的结构化知识库, 如 DBPedia<sup>[41]</sup>, Freebase<sup>[42]</sup>, YAGO<sup>[43]</sup>等, 提出了一个新的语义解析架构——查询图<sup>[44]</sup>. Yih 等人将自然语言问句转化成查询图的表达形式, 查询图可以直接被映射到逻辑表单, 这使得原本的语义解析问题被转变成了查询图的生成问题, 查询图的结构如图 5<sup>[44]</sup>所示, 它由四种类型的节点组成: 1. 存在实体, 指知识库中存在的实体, 图中用圆角矩形表示; 2. 存在变量, 它不是一个具体的实体, 图中用圆表示; 3. 聚合函数, 指对实体的修饰性语言等, 图中用菱形表示; 4.  $\lambda$  变量, 指最后输出的答案, 图中由阴影圆表示. 各个节点之间通过谓语关系链接. 而查询图的生成, 主要分为 3 步, 首先找到主题实体, 然后根据自然语言问句的谓语与知识库中关系的匹配度确定图的主链. 最后根据知识库中主链的其他附加条件为查询图添加聚合函数. 查找主题实体和确定主链是整个任务的关键, 因此整个问题转变成了一个阶段化的搜索问题. Yih 等使用了现有的实体链接系统<sup>[45]</sup>确定自然语言问句的候选主题实体, 构建查询图的第一节点. 主链的生成主要是通过知识库中搜索与主题实体关联的谓语关系, 根据知识库的关系链确定查询图的主链, 这是一个自然语言问句关系与知识库中谓语的匹配问题. 而两者匹配的模型与文献 [30] 的方法类似, 首先用卷积神经网络进行句子的语义表示, 在语义的表示上也是基于图 4 的 CNN 模型, 把自然语言问句的主题实体用特殊符号代替, 然后通过图 4 的卷积网络, 得到一个语义表示向量, 谓语关系链也通过一个 CNN 网络转

化成向量的表达形式,最后根据向量间的距离算法求自然语言问句与候选查询图主链的相似度,根据相似度确定查询图的主链.两个 CNN 网络的训练方法与 [30] 的训练方法类似,该实验的数据来自 WEBQUESTIONS,相比于之前的模型,该实验结果有了很大的提高,F1 值为 52.5%.相比于其他模型(如 Bordes 等人的模型<sup>[34]</sup>),该模型的主要优势如下:其一是在词向量初始化时,采用了同文献 [30] 的字母三元组的计数向量,这样不仅有助于控制输入空间的维度,而且能够包容输入单词中的一些轻微拼写问题.其二是该模型利用的卷积神经网络结构拥有更强大的语义表示能力.

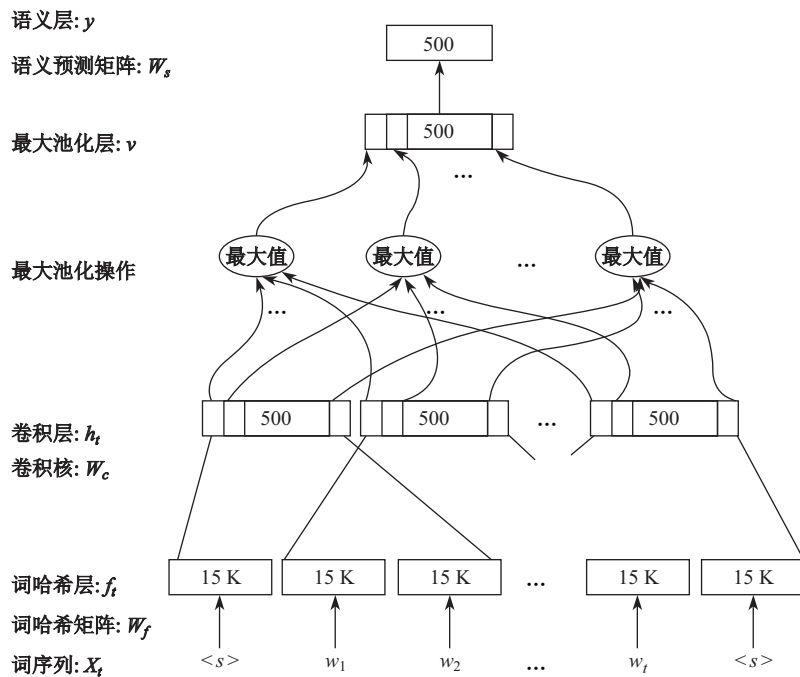


图4 CNNSM 模型

Fig. 4 CNNSM model

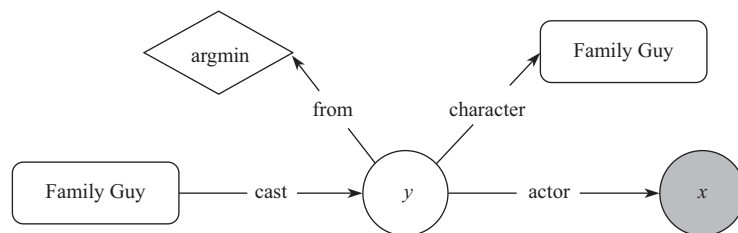


图5 问句“Who first voiced Meg on Family Guy”的查询图

Fig. 5 Query graph that represents the question “Who first voiced Meg on Family Guy”

与 Bordes 等人在文献 [34] 中提到的想法类似, Dong 等人<sup>[31]</sup>考虑从多个方面去分析问题,提取答案时,从答案的类型、答案的路径(即关系)、答案的内容三方面分析,并提出多栏卷积模型(MCCNN),如图 6<sup>[31]</sup>所示.问句端训练了 3 个不同参数的 CNN 模型,最后输出 3 个表示不同语义的向量表达式  $f_1(q)$ ,  $f_2(q)$ ,  $f_3(q)$ . 知识库中从答案路径(答案实体与问句实体之间的关系集合)、答案内容(直接与答案路径关联的实体和关系)、答案类型这 3 方面计

算答案端的向量表达式, 记为  $g_1(a)$ ,  $g_2(a)$ ,  $g_3(a)$ . 最后计算各个向量的相似度分数

$$s(q, a) = f_1(q)^T g_1(a) + f_2(q)^T g_2(a) + f_3(q)^T g_3(a).$$

该模型在 WEBQUESTIONS 上进行实验, F1 为 40.8%, 结果明显优于其他模型, 在其对比实验中, 其中 Berant 等人的方法<sup>[46]</sup>, Berant 与 Liang 的方法<sup>[47]</sup>, Bao 等人的方法<sup>[40]</sup>的 F1 值分别为 31.4%, 39.9%, 37.5%, 以上 3 种方法都基于语义解析, 且依赖于手工定义的特征和预定义的规则. 这说明这种卷积网络模型的效果与手工定义特征模型的效果相当, 甚至更优. 与 Bordes 等人<sup>[34]</sup>的实验对比表明, 多栏卷积模型也要优于 Bordes 等人<sup>[34]</sup>使用词嵌入和作为问题的表达方式的模型.

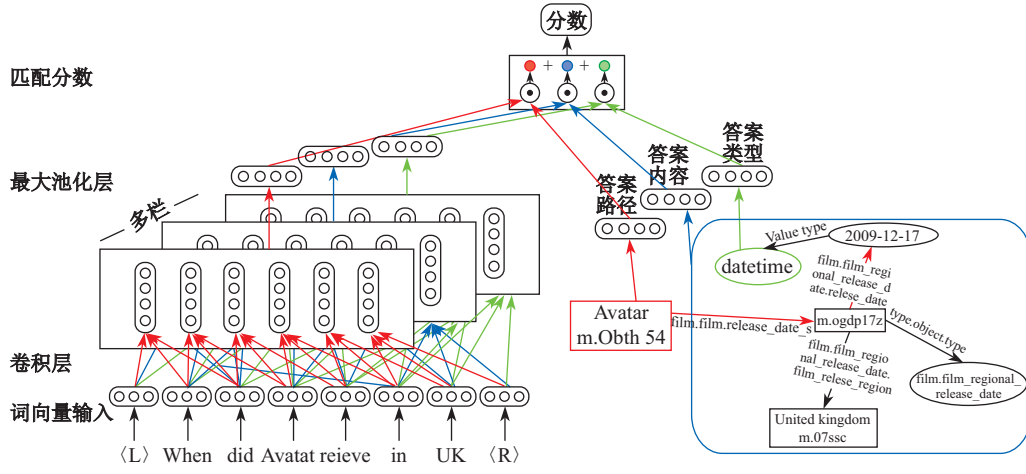


图6 MCCNN 模型

Fig. 6 MCCNN model

从以上几个代表性的工作来看, 基于卷积神经网络来处理面向知识库的问答系统时, 将传统的基于符号的语义分析的知识库问答看成基于语义表示的语义匹配学习任务, 提高了系统的准确率和召回率. 同时, 我们也可以看出, 如果采用通用的方法对自然语言问句或基于知识库的语义直接进行学习, 得到的效果往往不太好. 最好对自然语言问句进行多角度的分析, 学习多个语义表达式, 如从问句的主题、答案类型等方面与知识库中的实体、关系进行匹配.

### 3.2 基于卷积神经网络的文本问答

近年来, 对于卷积神经网络在面向文本的问答系统的应用, 研究人员主要关注于文本的抽取以及问答句的相似度计算的研究. 即给定一个自然语言问句, 通过一定的算法模型在自己的语料库或是在互联网上选出正确的答案.

Yu 等人<sup>[26]</sup>将问句与答案的匹配判断看成一个二分类问题, 并把每一个候选数据用一个三元组  $(q_i, a_{ij}, y_{ij})$  表示,  $y_{ij}$  的值为 0 或 1, 其中 1 表示相应的答案为正确答案, 0 表示错误答案. 最后的任务则是学习一个分类器去预测每一个三元组的  $y$  值, 即对于输入的一个问答对能够自动判断两者是否匹配,  $q$  和  $a$  是句子的向量表示, 在学习句子的向量表示时, Yu 等人采用的是二元词袋模型与卷积神经网络组合的方式, 即将二元词袋模型用于词向量的表示, 再通过只有一层卷积和一层池化的卷积神经网络来构建整个句子的语义表示. 根据 Bordes 等



人在文 [34] 中提出的问答句的相似度计算方法, 得到正确答案的概率为

$$p(y=1|\mathbf{q}, \mathbf{a}) = \sigma(\mathbf{q}^T \mathbf{M} \mathbf{a} + b),$$

其中  $\mathbf{q}^T \mathbf{M} \mathbf{a}$  是问句和答句两个向量的相似度计算分数,  $\mathbf{M}$  为相似度矩阵,  $\sigma$  为 sigmoid 激活函数, 使概率限制在 0~1 之间. 训练时, 为了解决词向量中没有专有名词和数字的缺陷, Yu 等人对问答句中共同出现的单词进行计数, 并将其计数值和这些单词的 TF-IDF 值与分布式词向量进行整合, 一起作为学习逻辑回归分类器的特征向量, 最后通过最小化交叉熵来训练模型, 训练的目标函数为

$$\ell = -\log \prod_n p(y_n | \mathbf{q}_n, \mathbf{a}_n) + \frac{\lambda}{2} \|\theta\|_F^2.$$

实验的数据集都来自 TREC<sup>[48]</sup>, 评估方法都采用 MAP(Mean Average Precision)标准和 MRR(Mean Reciprocal Rank)标准, Yu 等人<sup>[26]</sup>的 MAP 和 MPP 值分别为 0.711 3 和 0.784 6. 该结果较对比实验中的所有模型都要好, 之前的模型, 如文献 [48-51] 等中的方法都侧重于语法分析, 并不重视语义分析. 在之前的模型中, 文献 [52] 的模型表现较好, 它是唯一集中进行语义分析的模型. 在文献 [52] 的模型中, 先在问题和答案匹配对齐的单词中提取特征, 再将词特征进行聚合用以句子分类, 然后将这一组词特征与从外部资源获取的语义特征(如 WordNet, PILSA<sup>[53]</sup>等)结合. Yu 等人<sup>[26]</sup>的模型使用两种非分布特征——通过简单卷积神经网络构成的问答句匹配和以 IDF 值衡量的词匹配来进行语义语法分析, 该模型不依赖于外部的语义资源, 还可以对初始词向量进行训练.

Yu 等人<sup>[26]</sup>只是对整个模型进行概括, 与此想法类似, Severyn 等人<sup>[21]</sup>采用了相同的相似度计算方法, 但是 Severyn 等人的模型是基于多元词组模型. 该模型先利用 Word2Vec 工具将单词转化成词向量, 作为整个模型的初始词向量, 接着用 Skip-gram 模型训练, 之后, 再通过深层次的卷积网络构建问句与答案的向量表示, 并将句子向量看成是中间表达式的一部分. 具体的架构如图 7<sup>[21]</sup>所示, 经过卷积网络后, 分别得到两个句子的向量表达式, 再采用和文献 [26] 中同样的向量相似度计算方法, 计算两个向量的相似值, 这个相似值也作为中间表达式的一部分. 对于一些分布式词向量不能表示的专有名词和数字, 以及一些能够判断问题类型的词(如 what, when 等词), Severyn 等人<sup>[21]</sup>的处理方法与 Yu 等人<sup>[26]</sup>类似, 同样将这些信息作为额外的特征用于问答句之间的向量表达式的匹配度建立, 用特性向量  $\mathbf{x}_{\text{feat}}$  表示, 也作为中间表达式的一部分. 然后将所有中间向量输入到一个包括一个全连接层、隐藏层和 Softmax 层的神经网络, 其中全连接层将所有的中间向量进行连接, 隐藏层是用一个非线性函数作用于全连接后的输入向量, 最后, 通过 Softmax 输出, 概率分布公式为

$$p(y=j|\mathbf{x}) = \frac{e^{\mathbf{x}^T \theta_j}}{\sum_{k=1}^K e^{\mathbf{x}^T \theta_k}},$$

其中  $\theta_k$  为第  $k$  类的权值参数. 文献 [21] 的训练方法与文献 [26] 类似, 但是, 文献 [21] 采用端到端的形式训练, 文献 [26] 则是训练一个分类器.

文献 [21] 实验的数据集与评估标准与文献 [26] 一样, 这个模型的 MAP 和 MPP 值为 0.745 9 和 0.807 8, 相比文献 [26] 差不多有了 3% 的提高, 文献 [26] 使用了一个更有表现力的卷积神经网络来学习查询的中间表示和答案, 允许问答对之间进行更精确的匹配. 此外, 文献 [26] 的模型中还包括了问题和答案的中间表示, 这使得问答对之间有着更丰富的表示. 最

后, 文献 [26] 采用一种端对端的方式对系统训练, 而文献 [21] 则使用其深度学习系统作为逻辑回归分类器中的一个特性. 但是文献 [26] 模型的参数较多, 在实现上有一定的难度.

Hu 等人则提出了一种不一样的匹配思路, Hu 在文献 [54] 中给出了一种交互式的匹配模型(ARC-II), 其架构如图 8<sup>[54]</sup>所示. 该模型的基本思想是先把句子表达成为由句中单词的词向量的组合, 然后直接在两个句子的交互空间中构建 CNN 模型, 对匹配模式进行学习. 得到两个句子的整体交互的向量表示后, 最后通过一个神经网络(MLP<sup>[55]</sup>)来衡量这个匹配向量的匹配程度.

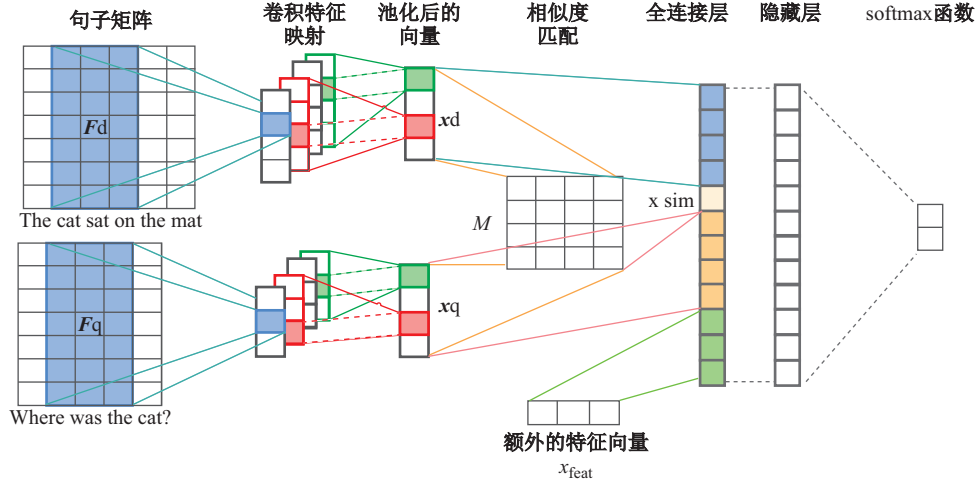


图 7 问答句匹配的深度学习架构

Fig. 7 Deep learning architecture for matching QA sentence

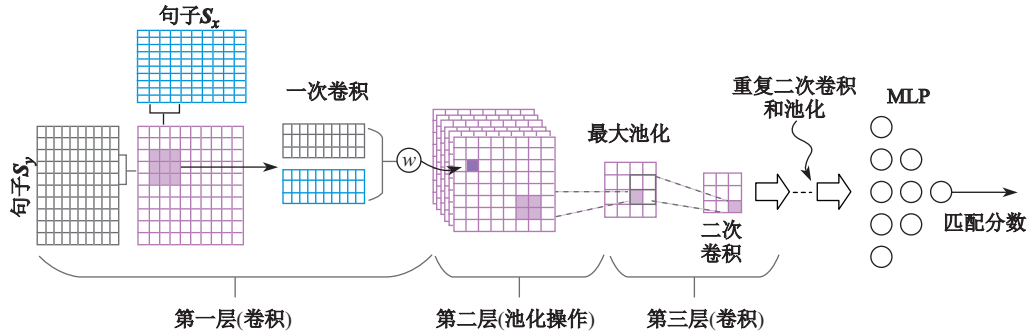


图 8 交互式语义匹配架构

Fig. 8 Architecture of interactive matching model

交互式的匹配结构卷积网络共有 3 层, 第一层为卷积层, 保留了问答句的局部信息, 利用卷积网络的滑动窗口, 对两个句子所有可能的词向量组合进行卷积操作, 得到两个句子交互的初步表示是一个矩阵, 称为特征映射矩阵. 经过第一层网络, 句子  $S_X$  从第  $i$  个单词开始组成的子序列与句子  $S_Y$  从第  $j$  个单词开始组成的子序列的一个值为

$$z_{i,j}^{(1,f)} \stackrel{\text{def}}{=} z_{i,j}^{(1,f)}(\mathbf{X}, \mathbf{Y}) = g\left(\hat{\mathbf{Z}}_{i,j}^{(0)}\right) \cdot \sigma\left(\mathbf{W}^{(1,f)} \hat{\mathbf{Z}}_{i,j}^{(0)} + b^{(1,f)}\right),$$

其中  $\hat{\mathbf{Z}}_{i,j}^{(0)}$  表示两个句子  $S_X$  与  $S_Y$  初始表达式的词向量的组合 ( $\hat{\mathbf{Z}}_{i,j}^{(0)} =$

$\left[ \mathbf{X}_{[i:i+k_1-1]}^T, \mathbf{Y}_{[j:j+k_1-1]}^T \right]^T$ ),  $\mathbf{W}^{(1,f)}$  表示第一层网络中第  $f$  个卷积核的权值,  $b^{(1,f)}$  是偏置项. 在该结构的第二层中, 对这个特征交互矩阵最大池化操作, 在每一个没有重叠的  $2 \times 2$  的子矩阵中选择最大值

$$z_{i,j}^{(2,f)} = \max \left( z_{2i-1,2j-1}^{(2,f)}, z_{2i-1,2j}^{(2,f)}, z_{2i,2j-1}^{(2,f)}, z_{2i,2j}^{(2,f)} \right).$$

进行池化操作后, 在结构的第三层, 再一次对降维后的矩阵进行卷积操作. 之后, 再进行一定次数的“池化+卷积”的操作, 得到两个句子交互的向量表示.

在训练该模型时, Hu 等人也采用了一种不同的训练方法, 通过采集大量三元组样本  $(x, y^+, y^-)$  使用基于排序的损失函数, 这旨在拉大正负样本的匹配分数的差距, 而不在意匹配度分数的绝对值大小:

$$e(x, y^+, y^-; \Theta) = \max \left( 0, 1 + s(x, y^-) - s(x, y^+) \right),$$

其中  $s(x, y^+)$  为正样本的匹配分数,  $s(x, y^-)$  为负样本的匹配分数, 训练过程中用随机梯度下降方法来不断优化模型. Hu 等人在句子的补充, 社交问答以及问句复述三种数据集上分别用 ARC-II 模型与 ARC-I 模型进行实验. 文献 [54] 中提到的 ARC-I 模型是一种并列式匹配模型, ARC-I 模型将两个句子分别通过卷积网络表示为两个定长的向量, 然后直接拼接两个向量并输入一个全连接的多层神经网络, 从神经网络的输出得到最终的匹配值, 实验结果显示交互式的匹配模型 ARC-II 要好于并列式的匹配模型 ARC-I. 而这两种模型都优于之前的一些模型. 相比于主题深度匹配模型 [56], Hu 等人 [54] 的利用卷积网络的深度匹配模型考虑了句子中词的顺序, 从而可以对两个句子的匹配关系进行相对完整的描述; 相比于树深度匹配模型 [57], Hu 等人 [54] 模型的学习框架更加灵活.

对于面向文本的自动问答, 问句和答案可以用同样的语义表示方法, 进行语义表示时, 词向量初始表示的选择对最后的结果有着重要的影响. 在进行语义匹配学习时, 基于语义学习的匹配模式是对句子的整体语义表示进行匹配, 对于某些不能用词向量表示的专有名词, 则需要单独学习. 而直接建模匹配模式能够考虑文本的一些局部性表达, 能有效捕捉问答句间短语的匹配.

### 3.3 困难与挑战

卷积神经网络在文本的语义表示以及文本的语义匹配上, 能够充分考虑句子的顺序, 且可以从多个方面考虑句子的语义, 在自动问答系统的应用上取得了较好的效果. 但与此同时, 在自动问答领域还有很多的挑战, 比如利用卷积网络, 只是对连续的短语进行组合, 只考虑了滑动窗口内的词序, 而无法捕捉远距离的依存关系, 也无法很好地表示复杂语义的句子, 对于能够捕捉句子复杂语义的模型, 其训练参数很多, 往往在实现上有一定难度. 目前, 大部分的问答系统都能处理简单的问答句子, 而对于一些复杂的或是不常见的问题的处理, 仍是当前问答领域的一大挑战. 特别是对面向知识库的问答系统, 除了需要对自然语言问句进行语义分析外, 还需依赖其他的技术, 如学习知识库语言的语义表示、实体链接技术等, 现有的模型在处理面向知识库的问答时, 一般比较好的效果也只在 50%~60% 区间.

## 4 总 结

本文主要对基于卷积神经网络下的自动问答技术进行综述, 针对卷积网络目前在面向知识库和面向文本的两大问答系统的应用, 重点关注卷积网络用于语义表示学习和语义匹

配两大技术点. 本文将问答匹配模型分为两类: 并列式匹配模型和交互式匹配模型, 并对两大知识库系统的不同实现技术进行展开. 知识库问答系统主要是自然语言问句与结构化语言的匹配问题, 面向文本的问答系统是自然语言问句和自然语言答案句的匹配问题. 在介绍基于卷积网络的两大问答系统的主流方法的同时, 对其存在的挑战和困难进行讨论. 自动问答是一个具有挑战性的研究领域, 特别是基于深度学习的自动问答系统是当下的研究热点之一. 相信随着深度学习的发展, 基于深度学习的自动问答技术会有更大的突破.

### [参 考 文 献]

- [1] KATZ B. Annotating the World Wide Web using natural language[C]//Proceedings of RIAO'97 Computer-Assisted Information Searching on Internet. 1997: 136-155.
- [2] SPINK A, GUNAR O. E-commerce web queries: eExcite and ask jeeves study[J/OL]. First Monday, 2001, 6(7). [2017-06-02]. [http://firstmonday.org/issues/issue6\\_7/spink/index.html](http://firstmonday.org/issues/issue6_7/spink/index.html).
- [3] ZHENG Z. AnswerBus question answering system[C]//Proceedings of the Second International Conference on Human Language Technology Research. Morgan Kaufmann Publishers Inc, 2002: 399-404.
- [4] 郑实福, 刘挺, 秦兵, 等. 自动问答综述[J]. 中文信息学报, 2002, 16(6): 46-52.
- [5] MOLLAD, VICEDO J L. Special section on restricted-domain question answering[J]. Computational Linguistics, 2006, 33(1): 41-61.
- [6] KWIATKOWSKI T, ZETTLEMOYER L, GOLDWATER S, et al. Lexical generalization in CCG grammar induction for semantic parsing[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2011: 1512-1523.
- [7] LIANG P, JORDAN M I, KLEIN D. Learning dependency-based compositional semantics[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (I). Association for Computational Linguistics, 2011: 590-599.
- [8] ZHANG Y, JIN R, ZHOU Z H. Understanding bag-of-words model: A statistical framework[J]. International Journal of Machine Learning and Cybernetics, 2010, 1(1/4): 43-52.
- [9] LANDAUER T K, FOLTZ P W, LAHAM D. An introduction to latent semantic analysis[J]. Discourse Processes, 1998, 25: 259-284.
- [10] BROWN P F, DESOUZA P V, MERCER R L, et al. Class-based  $n$ -gram models of natural language[J]. Computational Linguistics, 1992, 18(4): 467-479.
- [11] BENGIO Y, DUCHARME R, VINCENT P, et al. A neural probabilistic language model[J]. Journal of Machine Learning Research, 2003, 3: 1137-1155.
- [12] MIKOLOV T, YIH W T, ZWEIG G. Linguistic regularities in continuous space word representations[C]// Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics. 2013, 13: 746-751.
- [13] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space. [EB/OL]. [2017-06-02]. arXiv: 1301.3781. <https://arxiv.org/pdf/1301.3781.pdf>.
- [14] COLLOBERT R, WESTON J. A unified architecture for natural language processing: Deep neural networks with multitask learning[C]//Proceedings of the 25th International Conference on Machine Learning. ACM, 2008: 160-167.
- [15] MNIH A, HINTON G. Three new graphical models for statistical language modelling[C]//Proceedings of the 24th International Conference on Machine Learning. ACM, 2007: 641-648.
- [16] FREGE G. Funktion, Begriff, Bedeutung[M]. Gottingen: Vandenhoeck & Ruprecht, 2002.
- [17] HERMANN K M. Distributed representations for compositional semantics[D]. Oxford: University of Oxford, 2014.
- [18] 来斯惟. 基于神经网络的词和文档语义向量表示方法研究[D]. 北京: 中国科学院研究生院, 2016.
- [19] FUKUSHIMA K, MIYAKE S. Neocognitron: A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological Cybernetics, 1980, 36(4): 193-202.
- [20] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE. 1998, 86(11): 2278-2324.
- [21] SEVERYN A, MOSCHITTI A. Learning to rank short text pairs with convolutional deep neural networks[C]//Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2015: 373-382.

- [22] DAHL G E, SAINATH T N, HINTON G E. Improving deep neural networks for LVCSR using rectified linear units and dropout[C]// Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. 2013: 8609-8613.
- [23] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch[J]. *Journal of Machine Learning Research*, 2011, 12: 2493-2537.
- [24] BERGER A, LAFFERTY J. Information retrieval as statistical translation[C]//Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 1999: 222-229.
- [25] WU W, LU Z, LI H. Learning bilinear model for matching queries and documents[J]. *Journal of Machine Learning Research*, 2013, 14(1): 2519-2548.
- [26] YU L, HERMANN K M, BLUNSOM P, et al. Deep learning for answer sentence selection[EB/OL]. [2017-06-02]. arXiv: 1412.1632. <https://arxiv.org/pdf/1412.1632.pdf>.
- [27] SURDEANU M, CIARAMITA M, ZARAGOZA H. Learning to rank answers to non-factoid questions from web collections[J]. *Computational Linguistics*, 2011, 37(2): 351-383.
- [28] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]//Advances in Neural Information Processing Systems. 2013: 3111-3119.
- [29] 庞亮, 兰艳艳, 徐君, 等. 深度文本匹配综述[J]. *计算机学报*, 2017, 40(4): 985-1003.
- [30] YIH W, HE X D, MEEK C. Semantic parsing for single-relation question answering[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics. 2014: 643-648.
- [31] DONG L, WEI F, ZHOU M, et al. Question Answering over Freebase with Multi-Column Convolutional Neural Networks[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics. 2015: 260-269.
- [32] CAI Q, YATES A. Large-scale semantic parsing via schema matching and lexicon extension[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics. 2013: 423-433.
- [33] BORDES A, WESTON J, USUNIER N. Open question answering with weakly supervised embedding models [C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2014: 165-180.
- [34] BORDES A, CHOPRA S, WESTON J. Question answering with subgraph embeddings[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014: 615-620.
- [35] HUANG P S, HE X, GAO J, et al. Learning deep structured semantic models for web search using clickthrough data[C]//Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management. ACM, 2013: 2333-2338.
- [36] SHEN Y, HE X, GAO J, et al. Learning semantic representations using convolutional neural networks for web search[C]//Proceedings of the 23rd International Conference on World Wide Web. ACM, 2014: 373-374.
- [37] FADER A, ZETTLEMOYER L S, ETZIONI O. Paraphrase-driven learning for open question answering[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics. 2013: 1608-1618.
- [38] FADERA, SODERLAND S, ETZIONI O. Identifying relations for open information extraction[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2011: 1535-1545.
- [39] YAO X, VAN DURME B. Information extraction over structured data: Question answering with freebase[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics. 2014: 956-966.
- [40] BAO J, DUAN N, ZHOU M, et al. Knowledge-based question answering as machine translation[J]. *Cell*, 2014, 2(6): 967-976.
- [41] LEHMANN J, ISELE R, JAKOB M, et al. DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia[J]. *Semantic Web*, 2015, 6(2): 167-195.
- [42] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: A collaboratively created graph database for structuring human knowledge[C]//Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. ACM, 2008: 1247-1250.
- [43] SUCHANEK F M, KASNECI G, WEIKUM G. Yago: A core of semantic knowledge[C]//Proceedings of the 16th International Conference on World Wide Web. ACM, 2007: 697-706.
- [44] YIH S W, CHANG M W, HE X, et al. Semantic parsing via staged query graph generation: Question answering with knowledge base[C]// Proceedings of the Annual Meeting of the Association for Computational Linguistics. 2015: 1321-1331.
- [45] YANG Y, CHANG M W. S-MART: Novel tree-based structured learning algorithms applied to tweet entity linking[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. 2015: 504-513.
- [46] BERANT J, CHOU A, FROSTIG R, et al. Semantic parsing on freebase from question-answer pairs[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2013: 1533-1544.
- [47] BERANT J, LIANG P. Semantic parsing via paraphrasing[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics. 2014: 1415-1425.

- [48] WANG M, SMITH N A, MITAMURA T. What is the Jeopardy model? A quasi-synchronous grammar for QA[C]//Proceedings of EMNLP-CoNLL '07. 2007: 22-32.
- [49] HEILMAN M, SMITH N A. Tree edit models for recognizing textual entailments, paraphrases, and answers to questions[C]//Proceedings of the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. ACL, 2010: 1011-1019.
- [50] WANG M, MANNING C D. Probabilistic tree-edit models with structured latent variables for textual entailment and question answering[C]//Proceedings of the 23rd International Conference on Computational Linguistics. ACL, 2010: 1164-1172.
- [51] YAO X, VAN DURME B, CALLISON-BURCH C, et al. Answer extraction as sequence tagging with tree edit distance[C]//Proceedings of NAACL-HLT. 2013: 858-867.
- [52] YIH W, CHANG M W, MEEK C, et al. Question answering using enhanced lexical semantic models[C]//Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. ACL, 2013: 1744-1753.
- [53] YIH W, ZWEIG G, PLATT J C. Polarity inducing latent semantic analysis[C]//Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational Linguistics, 2012: 1212-1222.
- [54] HU B, LU Z, LI H, et al. Convolutional neural network architectures for matching natural language sentences[C]//Proceedings of Advances in Neural Information Processing Systems. 2015: 2042-2050.
- [55] BENGIO Y. Learning Deep Architectures for AI[M]. Foundations and Trends in Machine Learning. Boston, USA: Now Publishers Ins, 2009.
- [56] LU Z, LI H. A deep architecture for matching short texts[C]//Proceedings of Advances in Neural Information Processing Systems. 2013: 1367-1375.
- [57] WANG M, LU Z, LI H, et al. Syntax-based deep matching of short texts[C]//Proceedings of the 24th International Conference on Artificial Intelligence. AAAI Press, 2015: 1354-1361.

(责任编辑: 林 磊)

---

(上接第 65 页)

- [59] GLADKOVA A, DROZD A, MATSUOKA S. Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't[C]//HLT-NAACL, 2016: 8-15.
- [60] MIKOLOV T, YIH W, ZWEIG G. Linguistic regularities in continuous space word representations[C]//HLT-NAACL, 2013: 746-751.
- [61] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[C]//ICLR, 2015: 1-15.
- [62] GROVER A, LESKOVEC J. Node2vec: Scalable feature learning for networks[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016: 855-864.

(责任编辑: 李万会)